

TAGO et al Page 2 of 5 HIRA.0143

FIG.2

106

As seen in the example of text data relating to a gene, there are cases where the information symbolizing a particular theme is not a keyword. In the example of gene-related data, the ID or base sequence of a particular gene is also the information symbolizing a particular theme. In this example, even if the text data describes genes with identical base sequences, the groups of keywords that appear in that data could be very different sets.

## **AGCTAGCTAGCTAGCT**

PMID:P00001 PMID:P00005

201

202

## FIG.3

108

ID P00005

301

In a database search system in which databases with a great quantity of sets of text data are searched to extract and refer to text data that describes a desired theme, the searcher first classifies the entire text data into a plurality of groups in an arbitrary manner, and a cluster analysis is conducted on each of groups of keywords that are entered also in an arbitrary manner, in units of the aforementioned text data groups and in accordance with the frequency of appearance of the keywords. As a result, even in cases where the keywords appearing in the individual items of text data are different, it becomes possible to identify text data describing a desired theme as groups of text data.

302 .

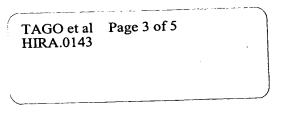


FIG.4

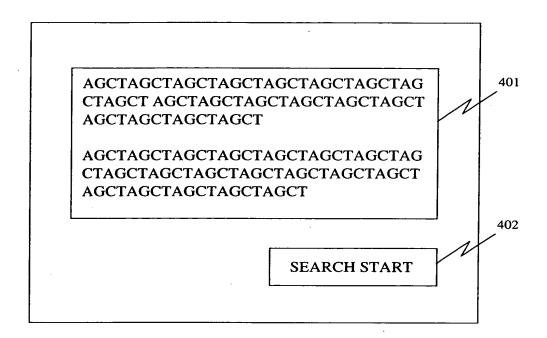


FIG.5

501	502	503
cell recognition	axon guidance	axon midline choice point recognition
		motor axon guidance
	axon extension	negative regulation of axon extension
		positive regulation of axon extension
cell-cell signaling	peptide hormone secretion	growth hormone secretion
		insulin secretion
	transmission of nerve impulse	action potential propagation
		regulation of action potential

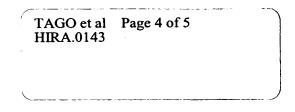
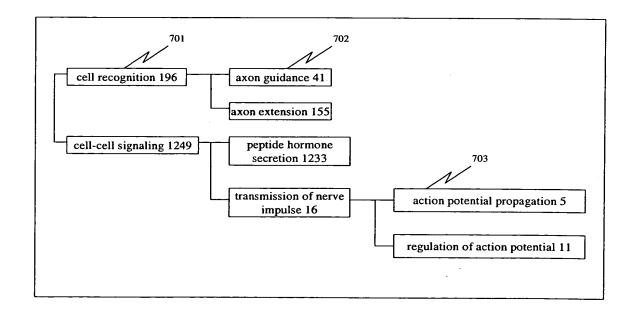


FIG.6

601	602	603
$\mathcal{N}$	N	$\mathcal{N}$
196	41	23
		18
	155	2
		153
1249	1233	39
		1194
	16	5
		11

FIG.7



TAGO et al Page 5 of 5 HIRA.0143

FIG.8

